**Data Set Information:**

The Data set is collected from UCI Machine learning repository archives ([https://archive.ics.uci.edu/ml/datasets/BLOGGER#](https://archive.ics.uci.edu/ml/datasets/BLOGGER)). The Data set is developed by collecting information to form a database and is done by questionnaires. This questionnaire is provided as oral, written and programming of a website which includes an internet questionnaire and the users can answer the questions as they wish. They entered their used websites, blogs and social networks during the day. After collecting questionnaires, the web addresses are gathered to get expected results. And finally, their truthfulness is checked by analysing their used web pages. As the results were the same, for getting a better and noiseless response, they will be put in the database.

consists of hundred instances among which sixty-eight instances are positive and the remaining instances are negative. Positive instances represent professional bloggers and negative instances indicate non-professional/ seasonal bloggers. There are five attributes in this dataset named Degree, Political Caprice, Topics, Local Media Turnover (LMT), Local, Political and Social Space (LPSS), and Professional Blogger (Pb) Described below.

1. **DEGREE:**

This attribute shows the education level of a blogger and has three values;

Low: for less educated people,

Medium: for B.Sc. level, and

High: for M.Sc. and Ph.D.

1. **CAPRICE:**

(Political Caprice) : It shows political affiliation of a blogger It has also three values;

LEFT: Bloggers affiliated with reformists party,

MIDDLE: bloggers who have no political affiliations, and

RIGHT: bloggers affiliated with a traditional conservative party.

1. **TOPICS:**

Bloggers are intended to blog in five areas named Impression, Political, Tourism, Scientific, and News.

1. **LMT:**

(Local Media Turnover) This attribute has two possible values: Yes/ No. “Yes,” value is used for those bloggers who believe in local media effect on blogging and vice-versa.

1. **LPSS :**

(Local, Political and Social Space) This attribute has also Yes/No values. “Yes,” value is used to denote those bloggers who believe that local, social and political conditions affect blogging and vice-versa.

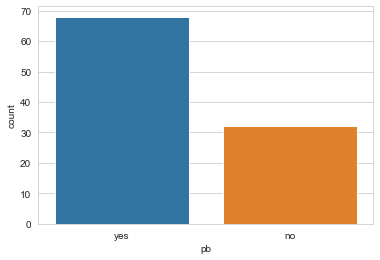
1. **PB:**

(Professional Blogger) It is a target class with Yes/No values. “Yes” value denotes a professional blogger, and “No” value indicates a temporary blogger.

1. **Data Exploration:**

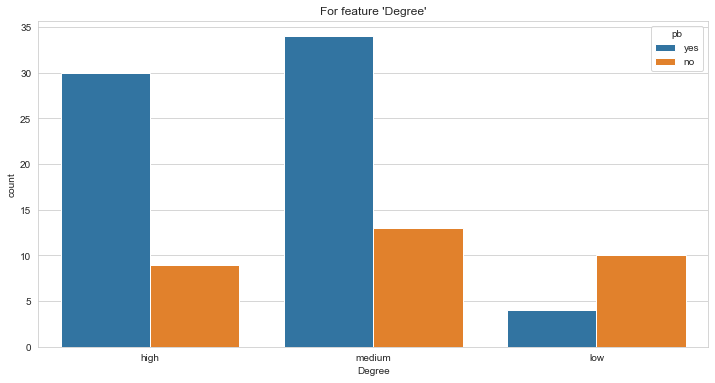
We have used Jupyter Note for pre-processing the data. From the Data set we identified the blogger classified into two groups: Professional Bloggers (Yes) and Temporary bloggers (No). Aim is to predict the blogger type.

***Figure 1****:*



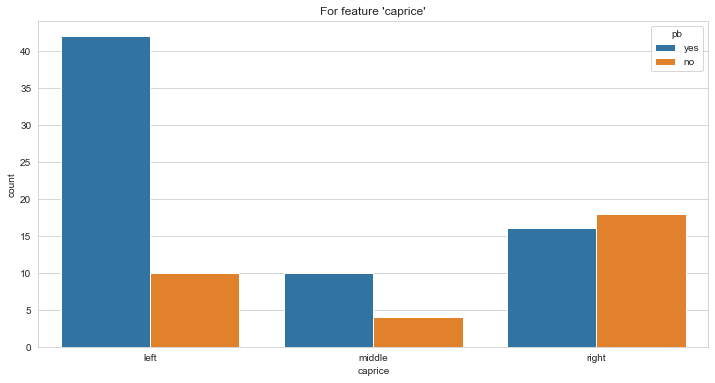
* Date set consists of hundreds of instances among which sixty-eight instances are positive and the remaining instances are negative.
* Positive instances represent professional bloggers and negative instances indicate non-professional/ seasonal bloggers

***Figure 2:***



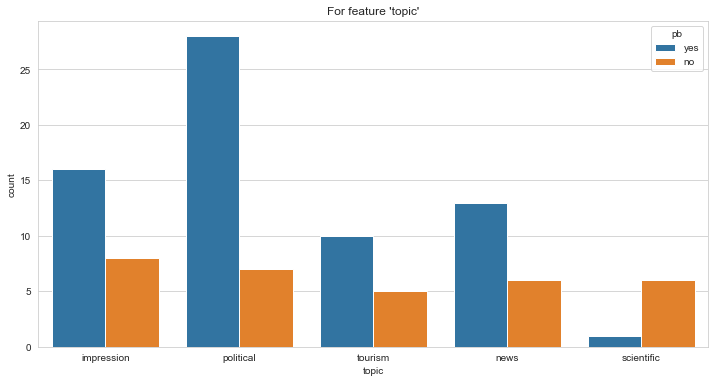
* Most Bloggers are higher degree and Ph.D. graduates
* Those with lower degree don’t have professional approach to blogging

***Figure 3:***



* Figure 3 interprets that most bloggers belong to the political party of leftists. This Party has a great tendency to political blogging.

***Figure 4:***

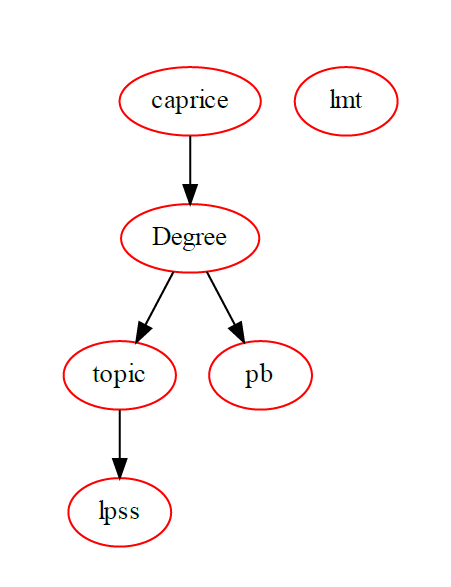


* Above picture depicts that most people who are interested in politics are professional bloggers.
* Impression, tourism and news are in the next place. Scientific bloggers are temporary bloggers.

## Question 1

**Bayesian Belief Network**

The Bayesian Belief network was created using the pomegranate Python Package, out of all the possible Bayseian networks that could be created, the below was chosen by the ‘exact’ algorithm.



### Steps to develop the Bayesian Belief network.

1.Model Structure

The model structure was learned from the data using the from\_samples method of the pomegranate package, we choose to use this as there are many possible structures and the best one is picked by the package.

2.CPT table entries

The fit method was called to calculate the CPT table entries.

3.Finally model.bake() was called to calculate the joint probabilities so that inference could be done.

**Methods used**-

1. model.probability() method was used to calculate the joint probability for a particular combination.
2. model.predict() method was used to calculate the MLE estimates of the missing values.

Eg:- model.predict([['high','middle','tourism','no','yes',None]]) => ‘YES’

**Resources used**:

<https://pomegranate.readthedocs.io/en/latest/>

<https://github.com/jmschrei/pomegranate/blob/master/tutorials>

**Installation Instructions:**

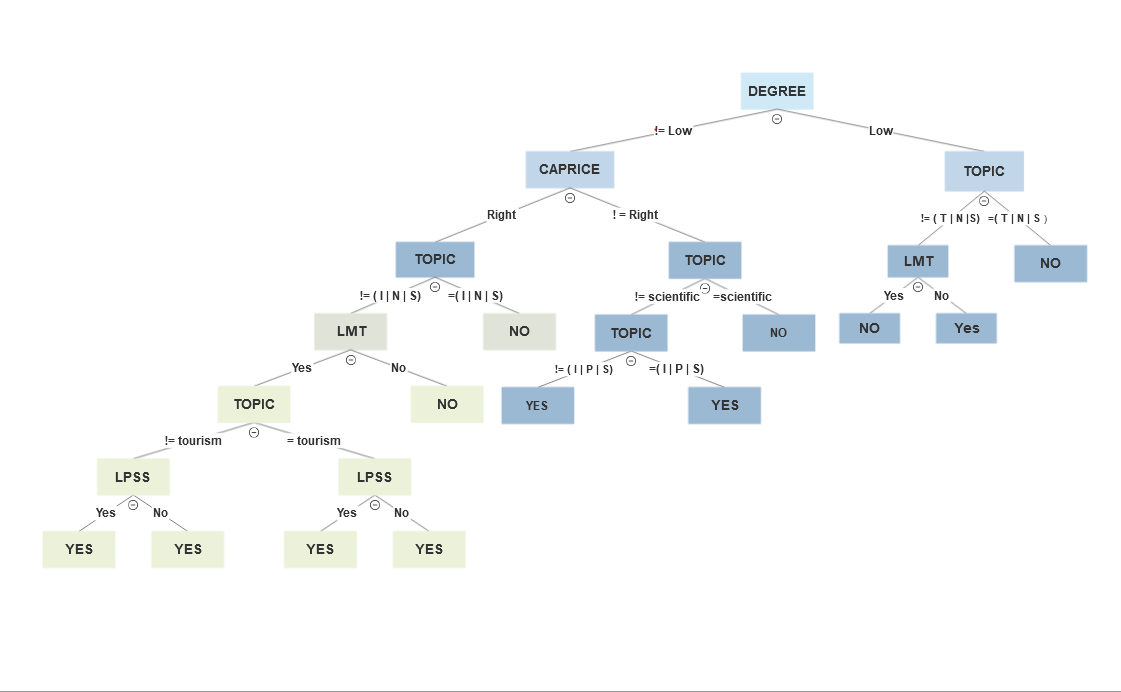
1. conda install pomegranate (required)
2. conda install -c alubbock graphviz pygraphviz (optional)
3. dot -c (optional)

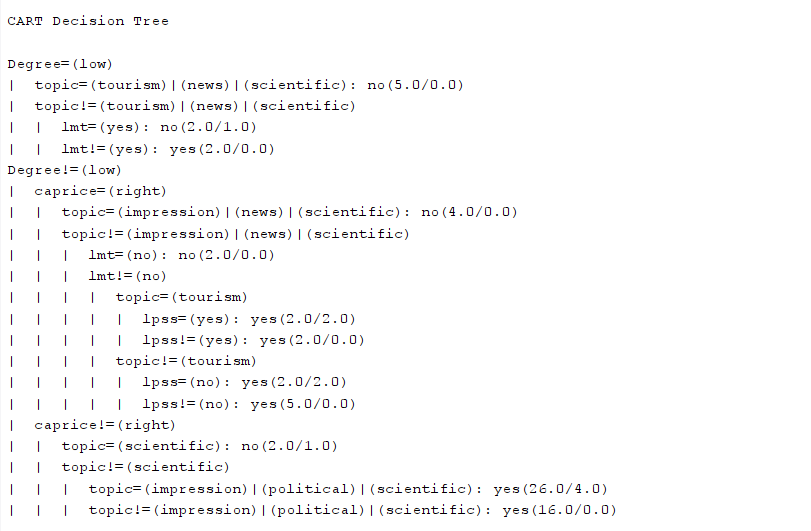
The optionally mentioned installation instructions are just for visualization that supports the pomegranate library to plot the bayesian networks and output it on a pdf.

## Question 2

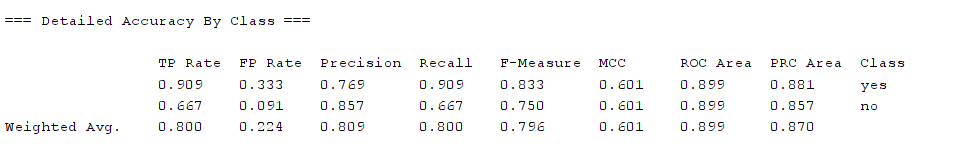
**Decision Tree algorithm:**

After preprocessing the data, we processed the data on Weka 3.8.5 and used the CART Decision tree algorithm. Then we obtained a Decision tree shown below





**Error and Precision rate of the above decision tree:**



**Prolog**

With the obtained Decision Tree, Prolog understands rules and facts. so we generated rules using Decision Tree and Implemented it.

**Instructions :**

1. Compile the .pl file
2. Give “start().” command.
3. This displays a set of instructions, please follow.

